blocks, processing, and other symbolic representations that directly or indirectly resemble the operations of data processing devices coupled to networks. These process descriptions and representations are typically used by those skilled in the art to most effectively convey the substance of their work to others skilled in the art. Numerous specific details are set forth to provide a thorough understanding of the present disclosure. However, it is understood to those skilled in the art that certain embodiments of the present disclosure can be practiced without certain, specific details. In other instances, well known methods, procedures, components, and circuitry have not been described in detail to avoid unnecessarily obscuring aspects of the embodiments. Accordingly, the scope of the present disclosure is defined by the appended claims rather than the forgoing description of embodiments.

[0272] When any of the appended claims are read to cover a purely software and/or firmware implementation, at least one of the elements in at least one example is hereby expressly defined to include a tangible, non-transitory medium such as a memory, DVD, CD, Blu-ray, and so on, storing the software and/or firmware.

1. A system comprising a network microphone device (NMD) of a media playback system, wherein the NMD comprises:

a microphone array;

a network interface;

at least one processor;

a tangible, computer-readable media;

instructions stored in the tangible, computer-readable media, wherein the instructions, when executed by the at least one processor, cause the NMD to perform functions comprising:

storing a data structure comprising nodes in a hierarchy representing the media playback system, wherein the data structure comprises (i) a root node representing the media playback system as a Home of the hierarchy, (ii) one or more first nodes in a first level, the first nodes representing respective devices of the media playback system as Sets of the hierarchy, and (iii) one or more second nodes in a second level as parents to one or more respective child first nodes to represent Sets in respective Rooms of the hierarchy, wherein the nodes in the hierarchy are assigned respective names;

recording, via the microphone array, audio into a buffer;

monitoring the recorded audio for keywords;

when a keyword is detected in the recorded audio, processing a voice input within a portion of the audio recording comprising the keyword, wherein processing the voice input comprises:

(i) determining, based on the data structure representing the media playback system, that one or more first keywords within the voice input represent respective target variables indicating one or more particular nodes of the data structure, each target variable referencing a name of a respective node of the data structure; and

(ii) determining that one or more second keywords within the voice input correspond to one or more playback commands; and

causing, via the network interface, one or more particular playback devices to play back audio content according to the one or more playback commands, wherein the one or more particular playback devices include (a) all playback devices represented by the one or more particular nodes of the data structure and (b) all playback devices represented by child nodes of the one or more particular nodes of the data structure.

2. The system of claim **1**, wherein determining that one or more first keywords within the voice input represent respective target variables comprises determining that at least one first voice keywords within the voice input represents a target variable referencing a name of a particular second node representing a particular Room, the particular Room including a first Set consisting of a first playback device and a second Set consisting of a second playback device, and wherein causing the one or more particular playback devices to play back audio content according to the one or more playback commands comprises causing the first playback device and the second playback device to play back the audio content in synchrony.

3. The system of claim **1**, wherein determining that one or more first keywords within the voice input represent respective target variables comprises determining that at least one first keyword within the voice input represents a target variable referencing a name of a particular first node representing a particular Set, the particular Set consisting of a first playback device and a second playback device in a bonded zone, and wherein causing the one or more particular playback devices to play back audio content according to the one or more playback commands comprises causing the first playback device and the second playback device to play back respective channels of the audio content in synchrony.

4. The system of claim **1**,

wherein the data structure further comprises one or more third nodes in a third level as parents to one or more respective child second nodes to represent Rooms in respective Areas of the hierarchy,

wherein determining that the one or more first keywords within the voice input represent respective target variables comprises determining that at least one first voice keyword within the voice input represents a target variable referencing a name of a particular third node representing an Area including a first Room and a second Room, the first Room including a first Set that consists of a first playback device and the second Room including a second Set that consists of a second playback device, and

wherein causing the one or more particular playback devices to play back audio content according to the one or more playback commands comprises causing the first playback device and the second playback device to play back the audio content in synchrony.

5. The system of claim **4**, wherein causing the one or more particular playback devices to play back audio content according to the one or more playback commands comprises causing the first playback device and the second playback device to form a synchrony group.

6. The system of claim **1**, wherein determining that one or more first keywords within the voice input represent respective target variables comprises determining that at least one first voice command within the voice input represents a target variable referencing a name of the root node, and wherein causing the one or more particular playback devices to play back audio content according to the one or more